Tararira: Music retrieval by sung query

Ernesto López Martín Rocamora
elopez@fing.edu.uy rocamora@fing.edu.uy

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería de la Universidad de la República
Uruguay

10th Brazilian Symposium on Computer Music, 2005

Abstract

The problem of music retrieval by sung query consists of building a machine capable of simulating the cognitive process of identifying a musical piece from a few sung notes of its melody. In this talk, the algorithms of pitch tracking, onset detection and melody matching used in the system Tararira are described. Much effort has been put on automatic transcription of singing voice as it is a key factor in the overall performance. A novel way of combining note by note matching with a recent approach based on pitch time series matching is introduced.

Outline

Outline

Outline

Outline

Multimedia content access
Query by humming systems

- Melody - memorable and representative in Western music.
- Query By Humming - practical and efficient way of access.
- Cognitive processes - very hard to simulate on a machine.

Multimedia content access
Query by humming systems

- Melody - memorable and representative in Western music.
- Query By Humming - practical and efficient way of access.
- Cognitive processes - very hard to simulate on a machine.

Multimedia content access
Query by humming systems

- Melody - memorable and representative in Western music.
- Query By Humming - practical and efficient way of access.
- Cognitive processes - very hard to simulate on a machine.

Multimedia content access
Query by humming systems

- Melody - memorable and representative in Western music.
- Query By Humming - practical and efficient way of access.
- Cognitive processes - very hard to simulate on a machine.

Outline

Previous Solutions Proposed
Database and matching approaches

Database

Music in symbolic notation (e.g. MIDI)

Matching approaches

- Note Sequence Matching (traditional approach)
- Pitch Time Series Matching (recent approach)

Note Sequence Matching

Method

- Query trancription to note sequence
- Search of best occurrences of note pattern

Drawbacks

- Automatic transcription errors reduce performance

Note Sequence Matching

Method

- Query trancription to note sequence
- Search of best occurrences of note pattern

Drawbacks

- Automatic transcription errors reduce performance

Pitch time series matching

Method

- Query fundamental frecuency (F0) estimation
- Comparison with melodies codified as pitch time series

Drawbacks

- Computational time becomes prohibitive
- Query must be a previously define melody fragment

Pitch time series matching

Method

- Query fundamental frecuency (F0) estimation
- Comparison with melodies codified as pitch time series

Drawbacks

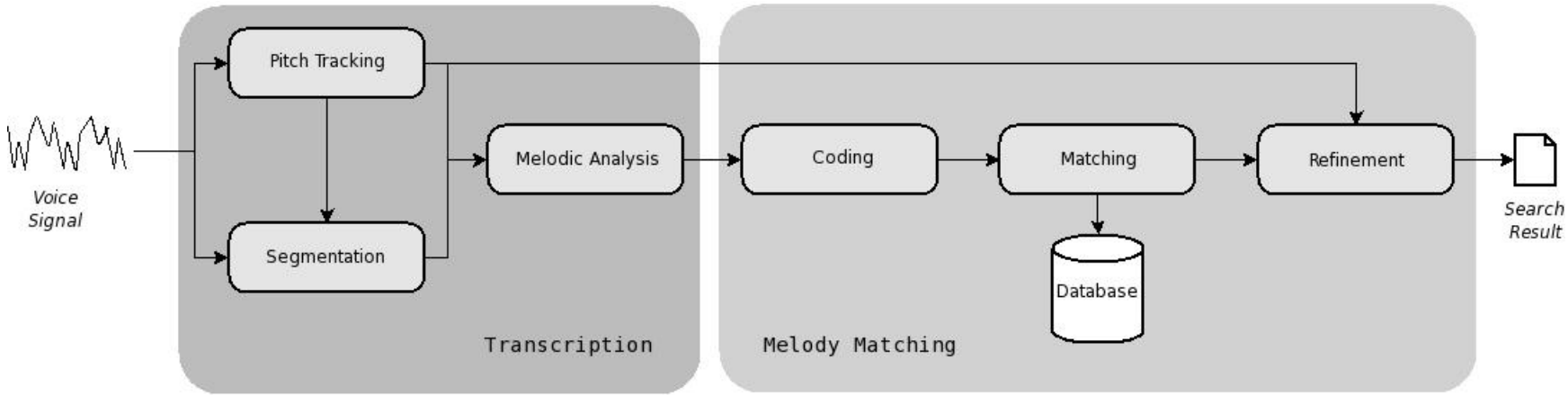
- Computational time becomes prohibitive
- Query must be a previously define melody fragment

Outline

Tararira System Overview

Transcription

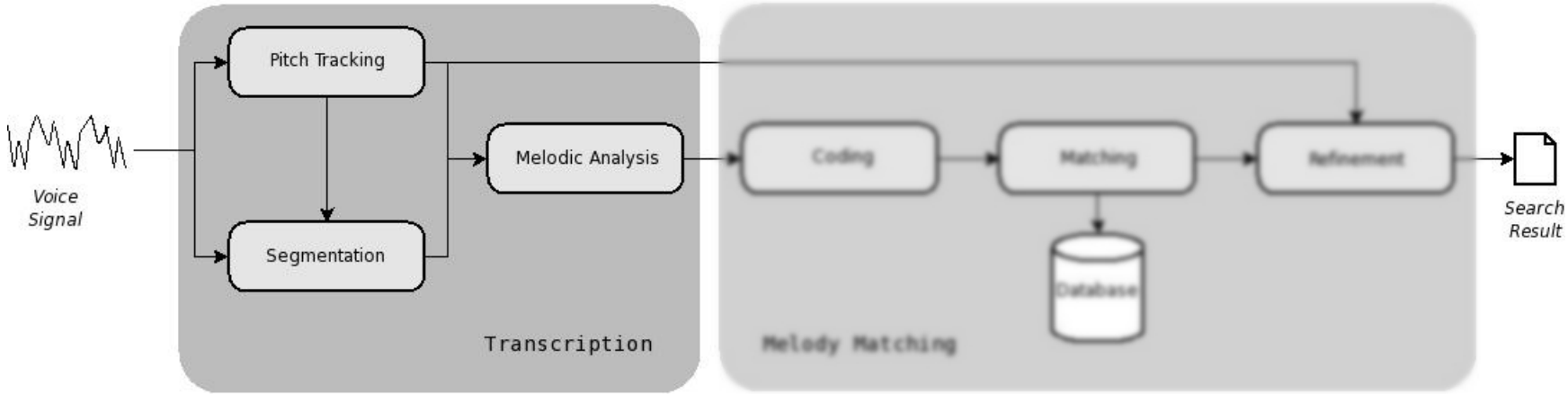
- Pitch Tracking to establish notes pitch
- Audio Segmentation to determine note boundaries
- Melodic Analysis to adjust pitches to tempered scale



Tararira System Overview

Transcription

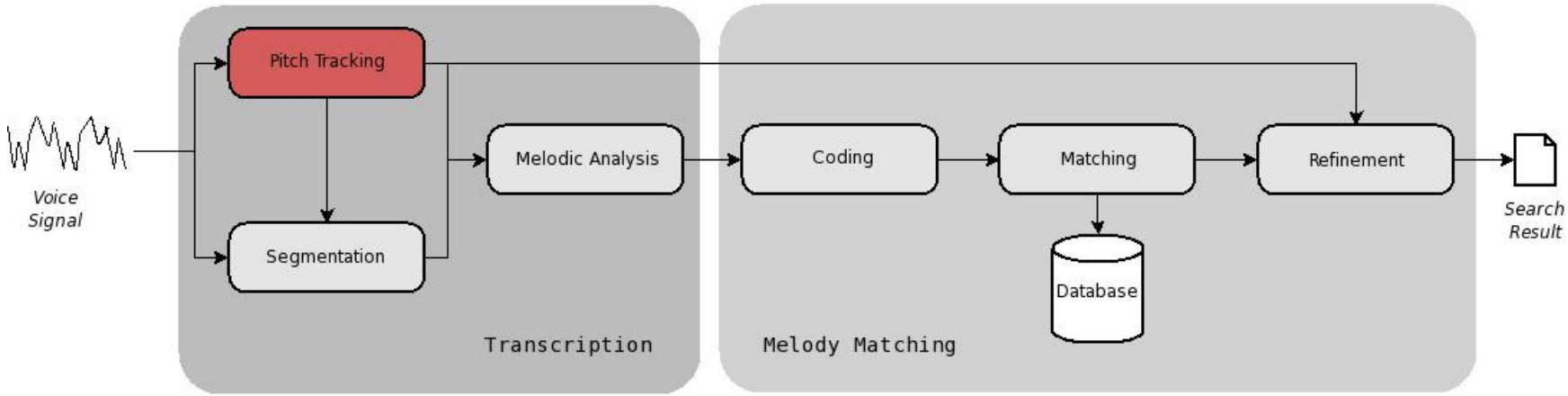
- Pitch Tracking to establish notes pitch
- Audio Segmentation to determine note boundaries
- Melodic Analysis to adjust pitches to tempered scale



Tararira System Overview

Transcription

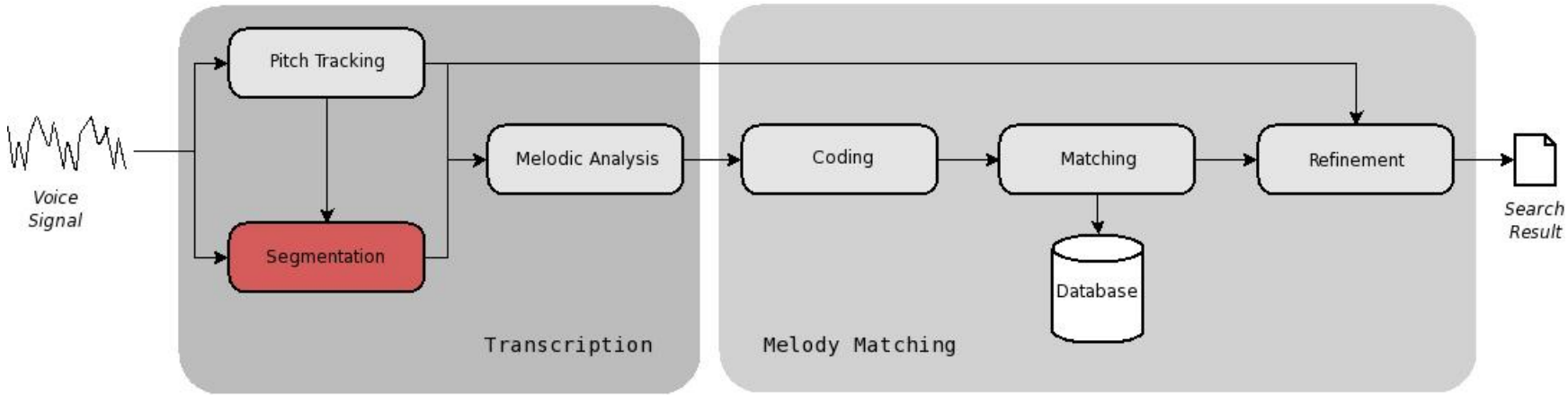
- Pitch Tracking to establish notes pitch
- Audio Segmentation to determine note boundaries
- Melodic Analysis to adjust pitches to tempered scale



Tararira System Overview

Transcription

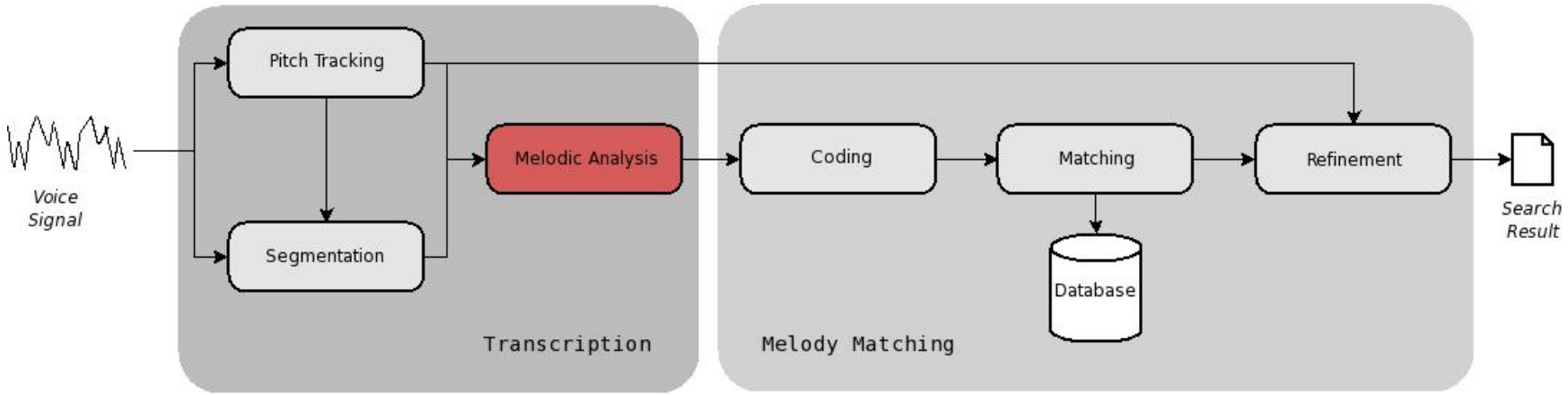
- Pitch Tracking to establish notes pitch
- Audio Segmentation to determine note boundaries
- Melodic Analysis to adjust pitches to tempered scale



Tararira System Overview

Transcription

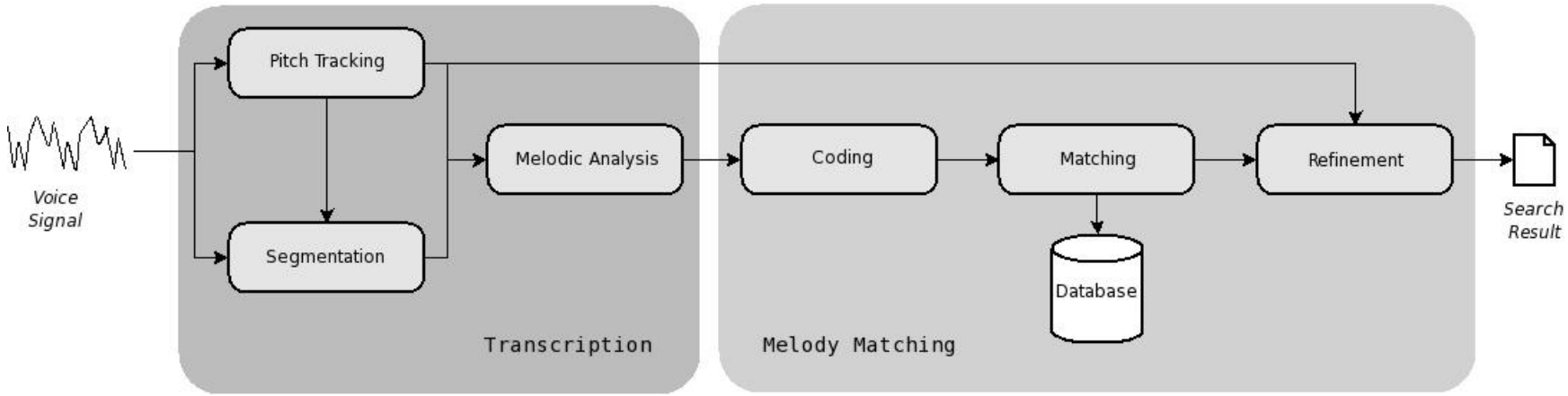
- Pitch Tracking to establish notes pitch
- Audio Segmentation to determine note boundaries
- Melodic Analysis to adjust pitches to tempered scale



Tararira System Overview

Melody Matching

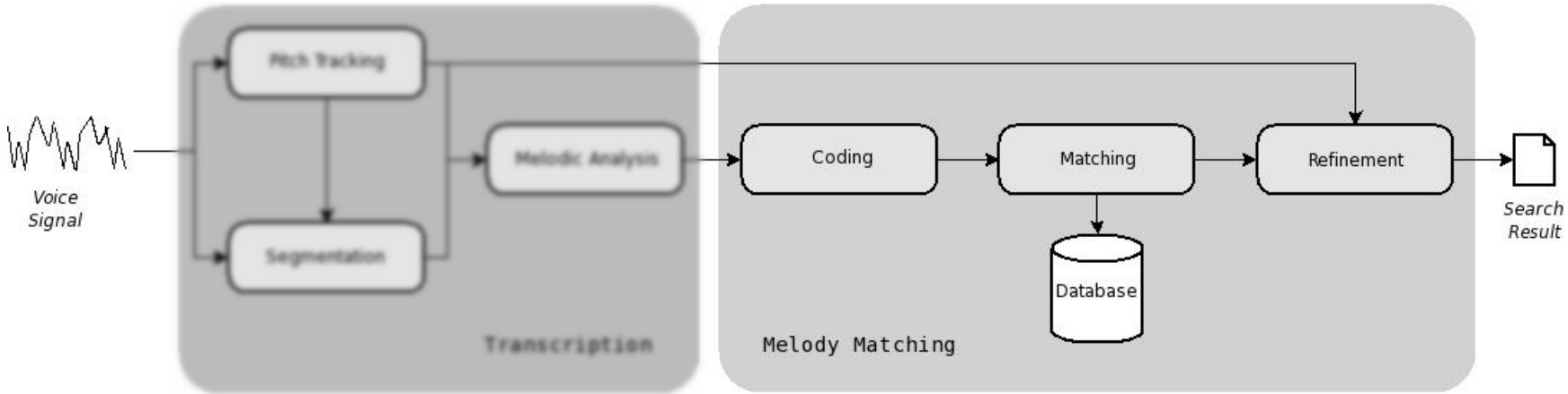
- Note sequence coding for pitch and tempo invariance
- Find good occurences considering flexible similarity rules
- Selection refinement using pitch time series



Tararira System Overview

Melody Matching

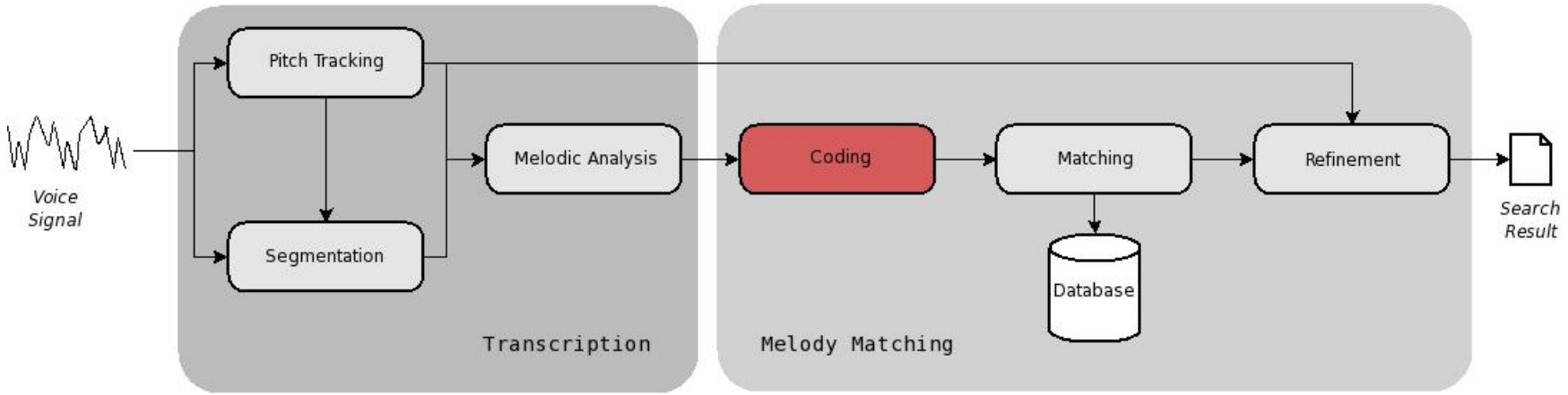
- Note sequence coding for pitch and tempo invariance
- Find good occurences considering flexible similarity rules
- Selection refinement using pitch time series



Tararira System Overview

Melody Matching

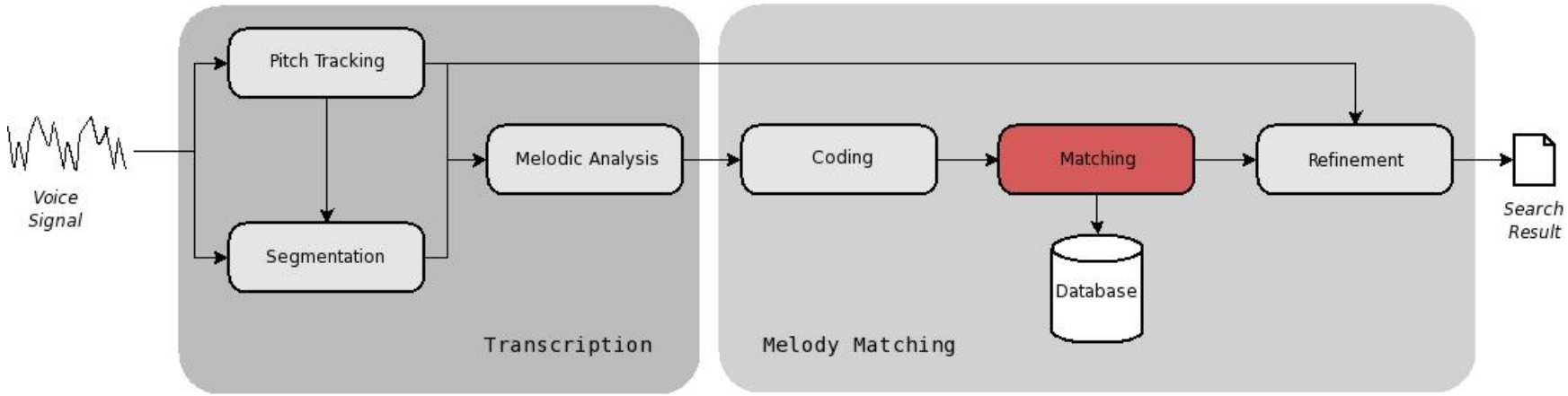
- Note sequence coding for pitch and tempo invariance
- Find good occurences considering flexible similarity rules
- Selection refinement using pitch time series



Tararira System Overview

Melody Matching

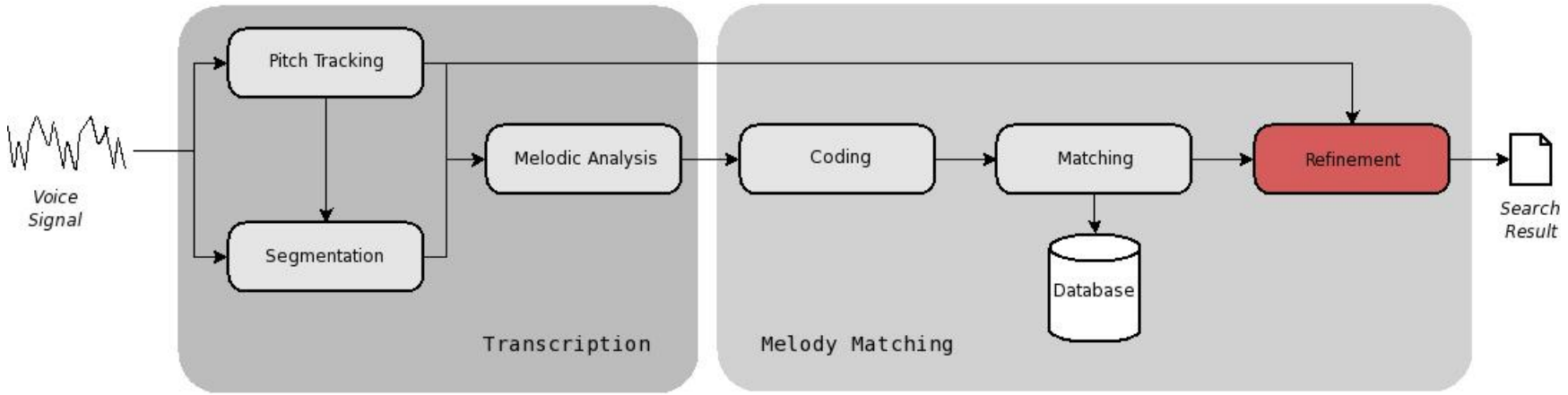
- Note sequence coding for pitch and tempo invariance
- Find good occurences considering flexible similarity rules
- Selection refinement using pitch time series



Tararira System Overview

Melody Matching

- Note sequence coding for pitch and tempo invariance
- Find good occurences considering flexible similarity rules
- Selection refinement using pitch time series



Outline

Automatic Voice Transcription

Automatic transcription goal

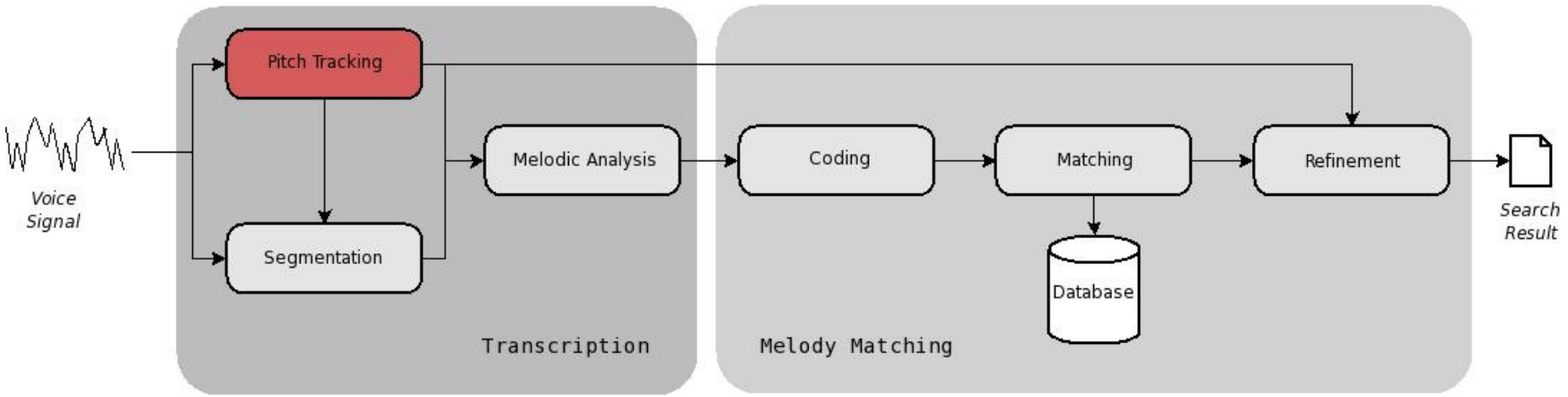
- Get the sequence of notes that best represent the query
- Each note characterized by pitch, onset time and duration

Problems

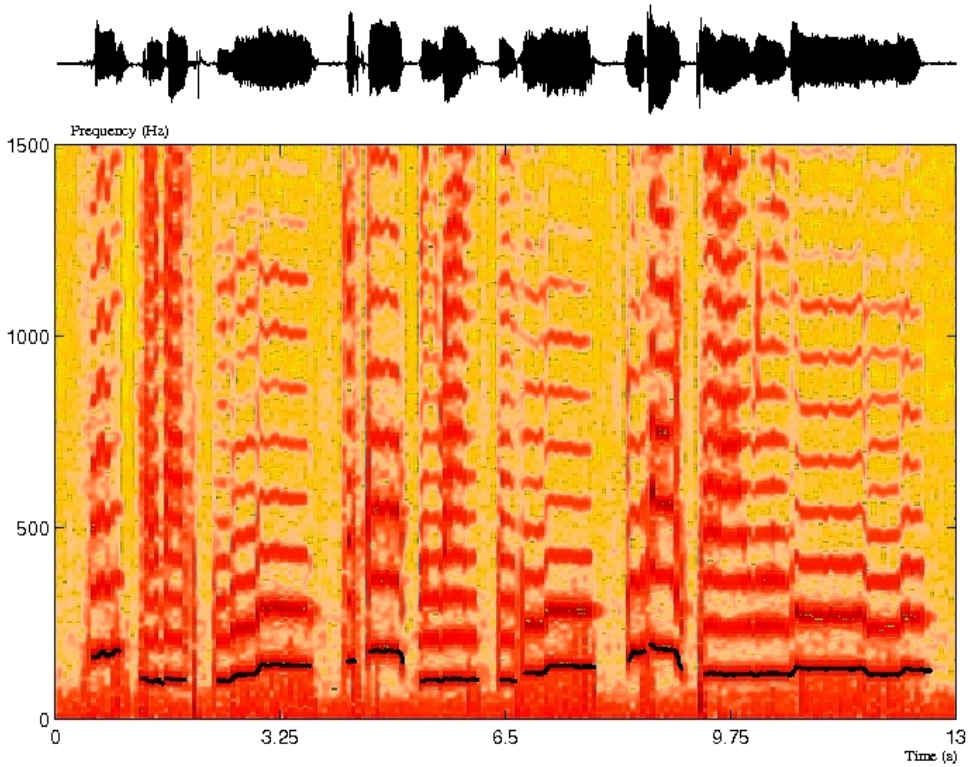
- Singing voice is the most difficult instrument to deal with
- Query not tuned to the equal tempered scale

Pitch Tracking Algorithm

Estimate fundamental frequency to establish notes pitch.



Pitch Tracking Algorithm

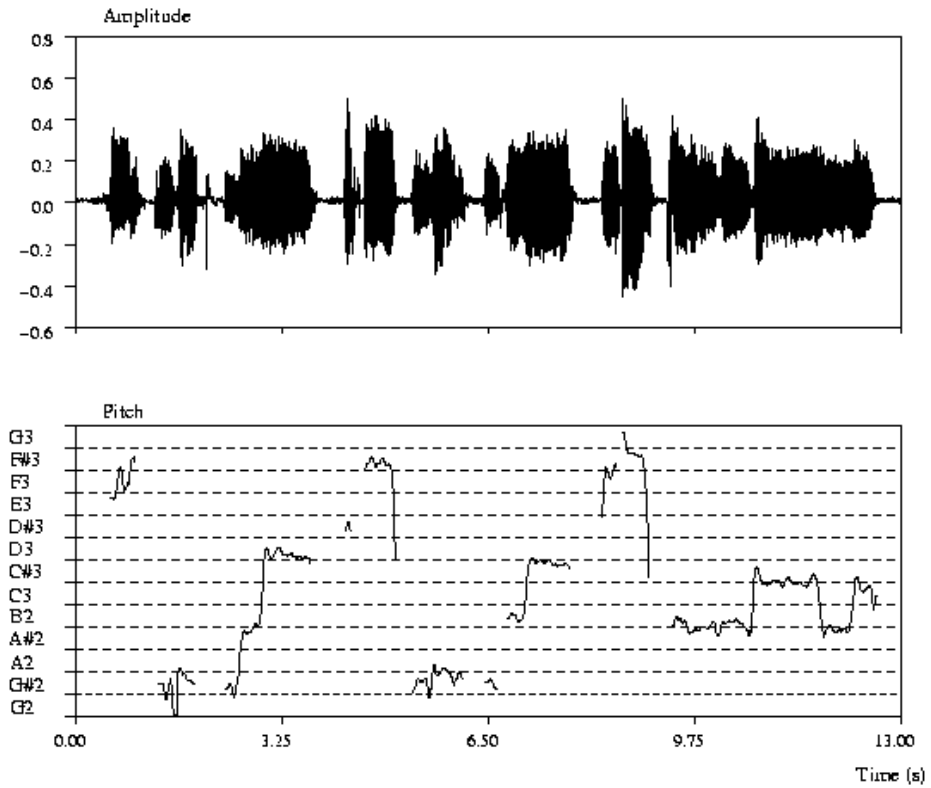


Difference Function

$$d(\tau) = \sum_{j=1}^{W} (x_j - x_{j+\tau})^2$$

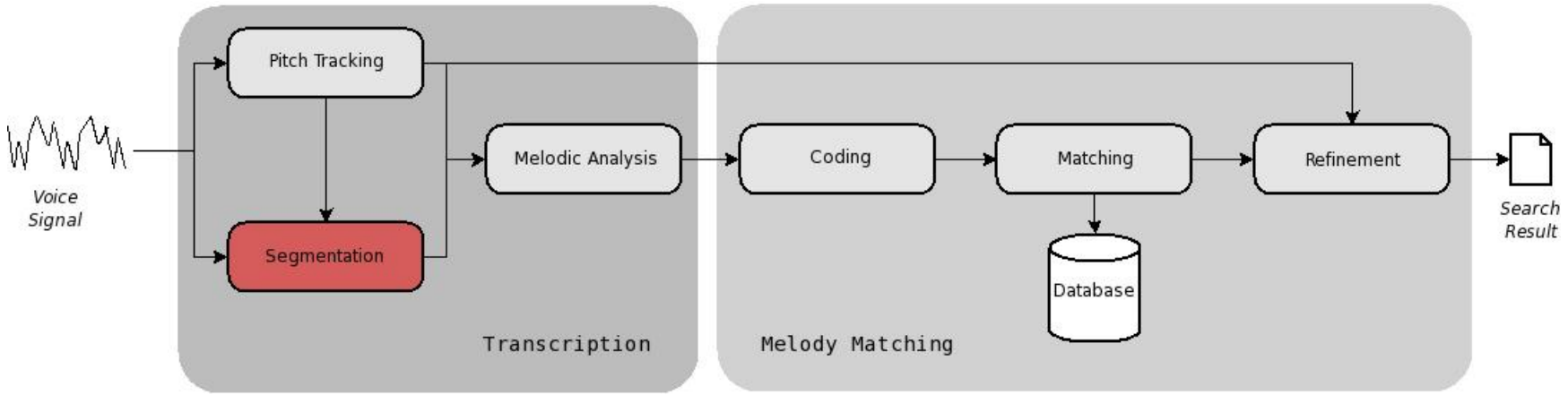
Pitch Tracking Algorithm

Fundamental Frequency Contour



Audio Segmentation

Establish onset times and durations.



Audio Segmentation

Singing voice note boundaries

- hard attack - sudden energy increment
- soft attack - gradual energy increase
- no attack - pitch changes without energy increase

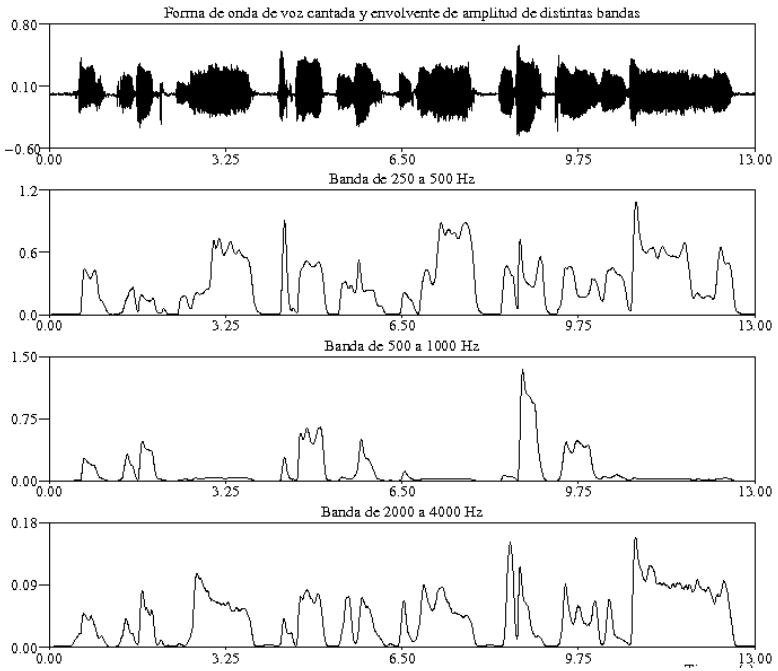
Algorithm

Look for signs of events in:

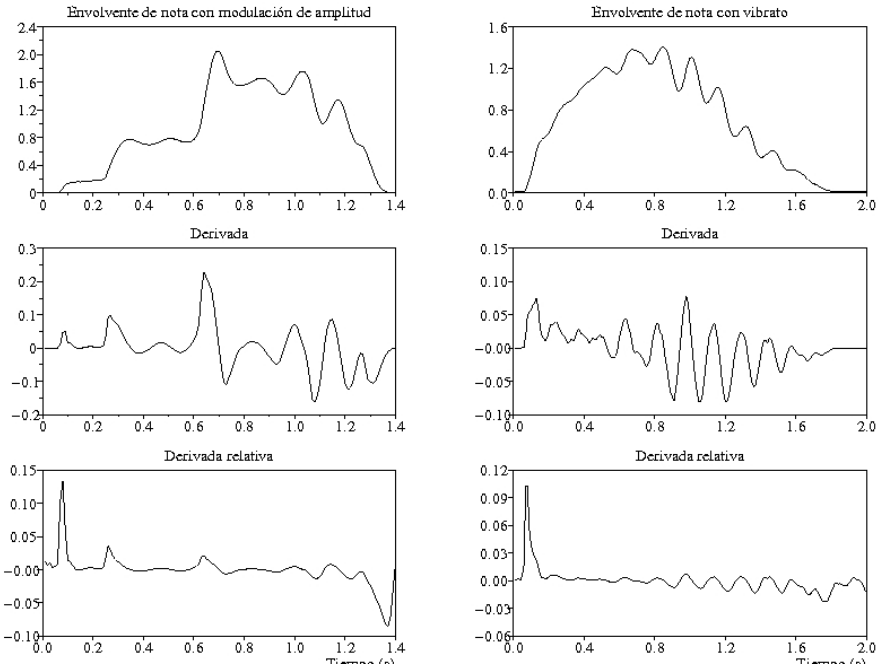
- amplitude envelope
- fundamental frequency contour

Energy changes

Sub-band Envelope Analysis

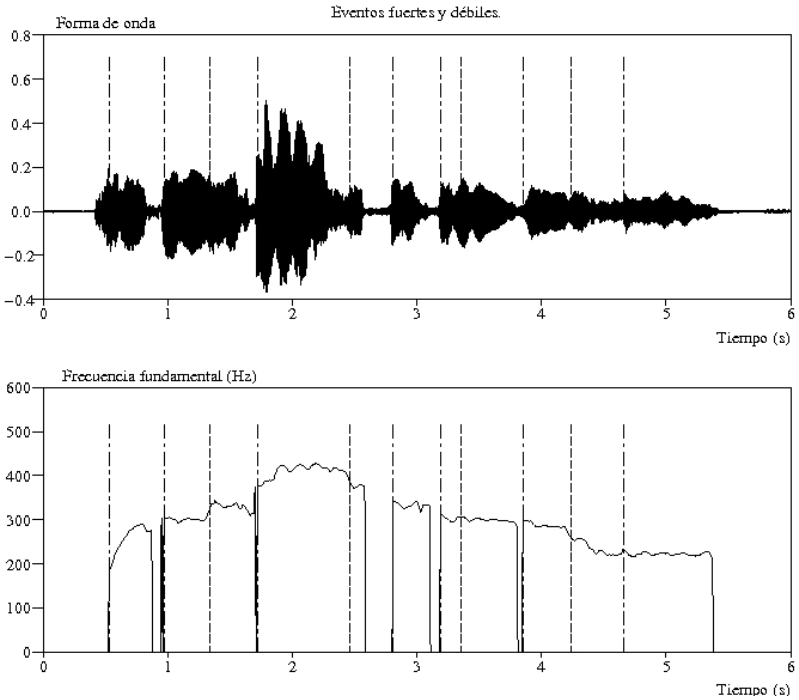


Envelope Derivatives

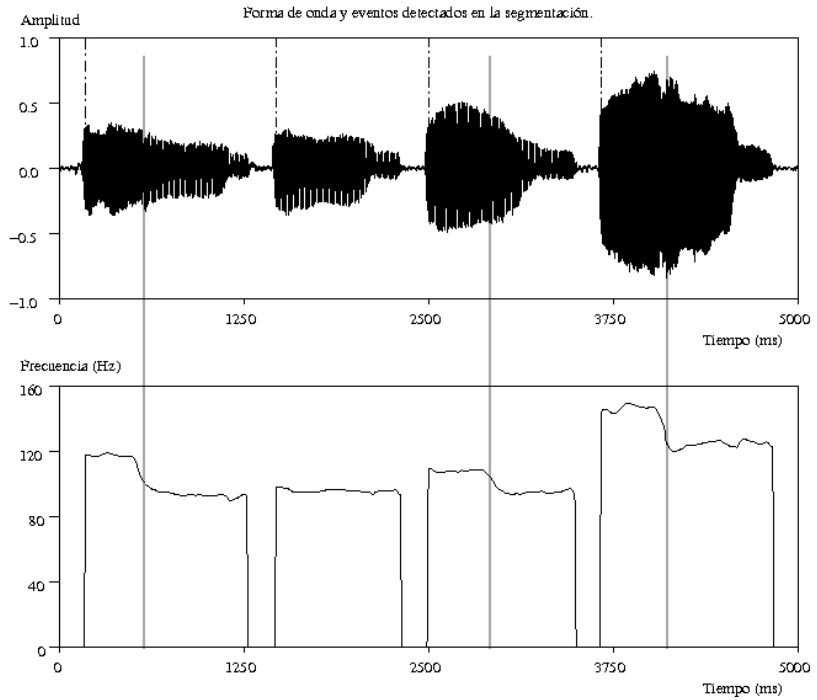


Pitch Changes

Weak events validation

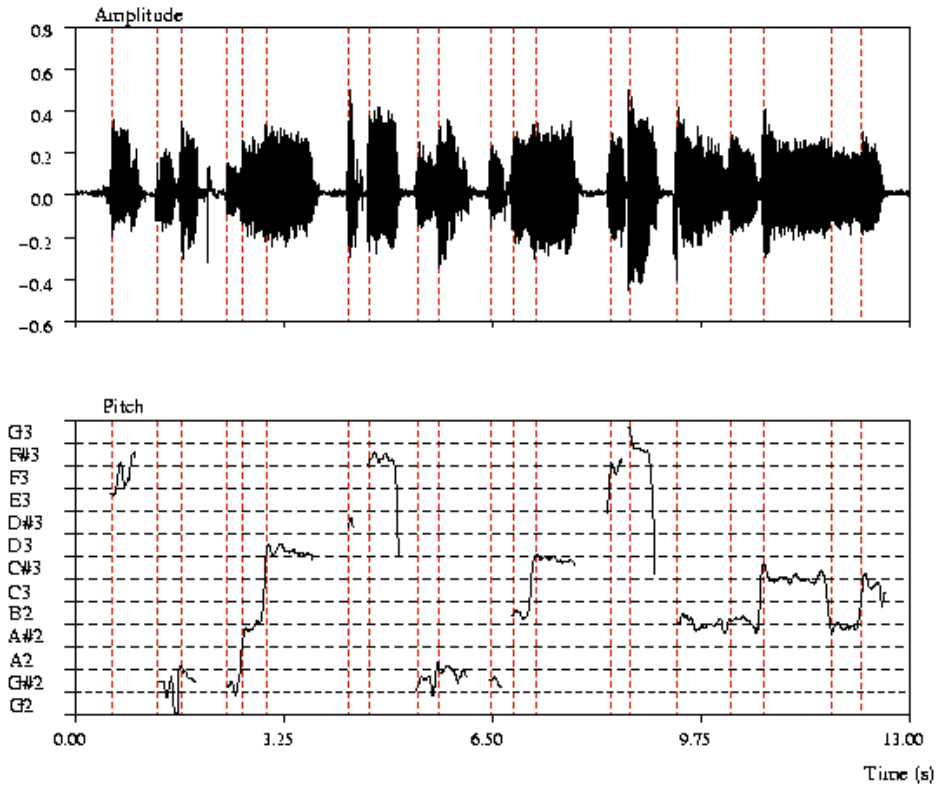


Evident pitch changes



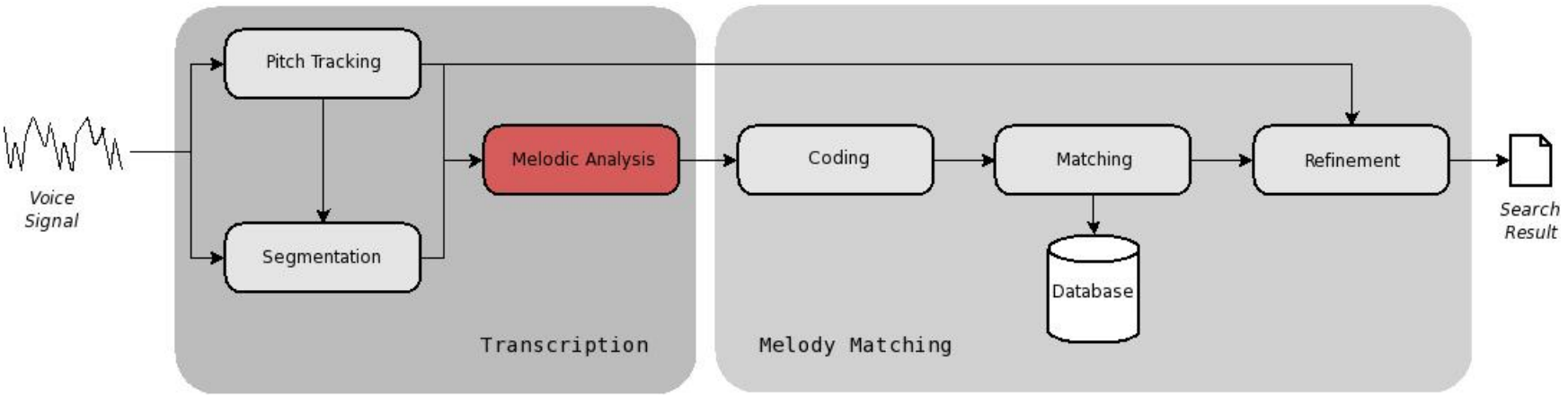
Audio Segmentation

Note Onsets



Melodic Analysis

Adjust note pitches to the equal tempered scale.



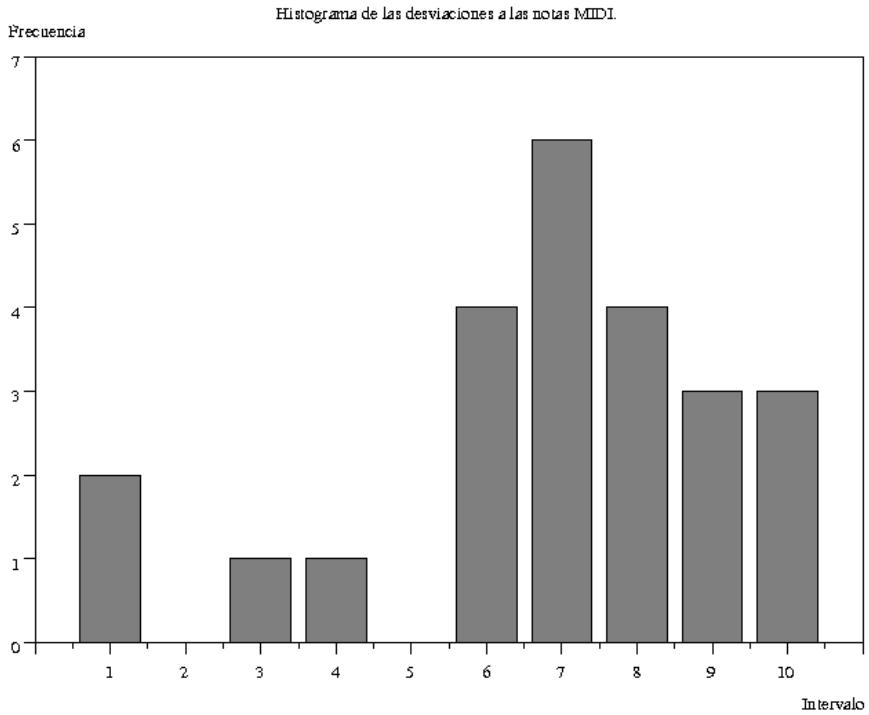
Melodic Analysis

Problem

Query does not respect equal tempered scale reference and intervals

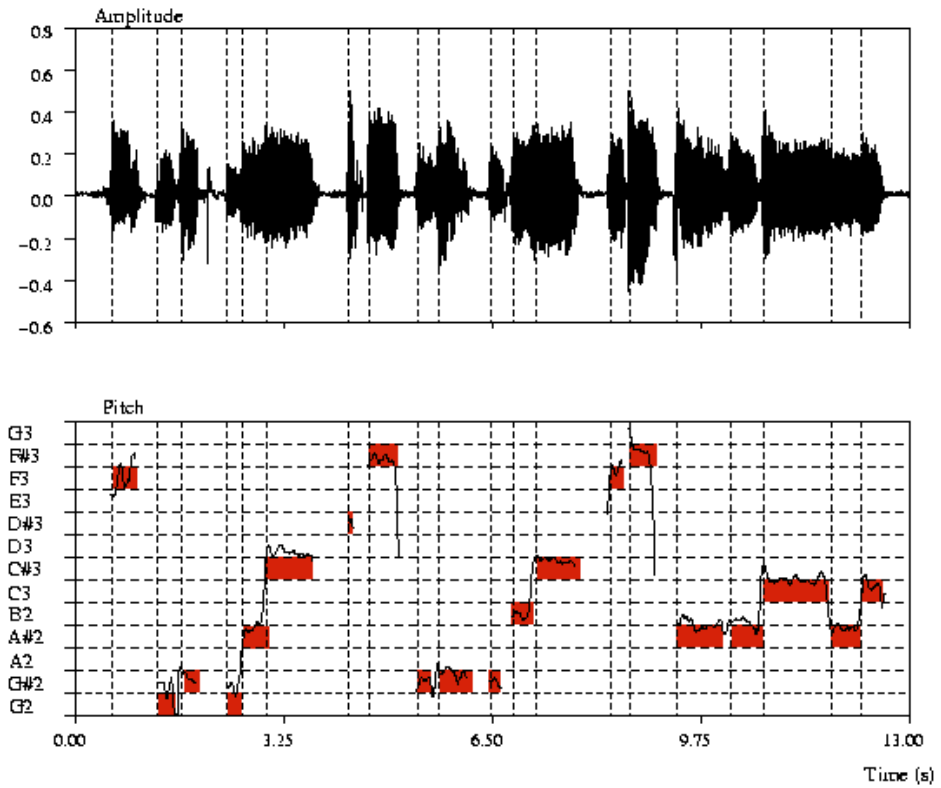
Adjustment method

- Hypotesis: reference tone held in mind
- Reference tone estimation by most frequent deviation



Melodic Analysis

Query Transcription



Outline

Melody Matching

Requirements

A melody can be identified despite being performed:

- at different pitch
- at different tempo
- with sporadic modifications or errors

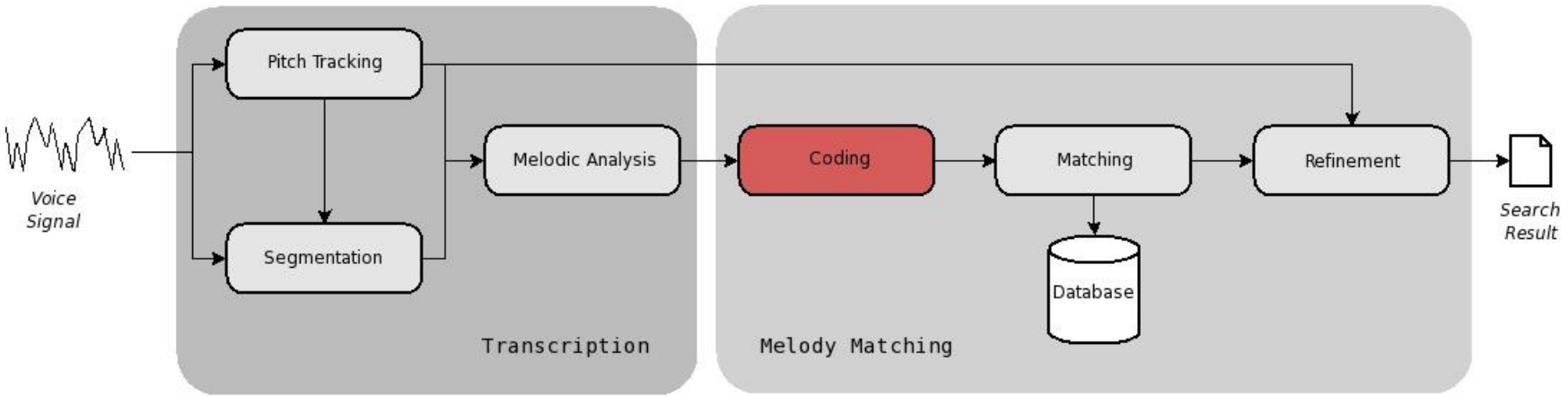
Algorithm

The algorithm provides:

- pitch and tempo invariance by note encoding
- error tolerance by flexible similarity rules

Coding

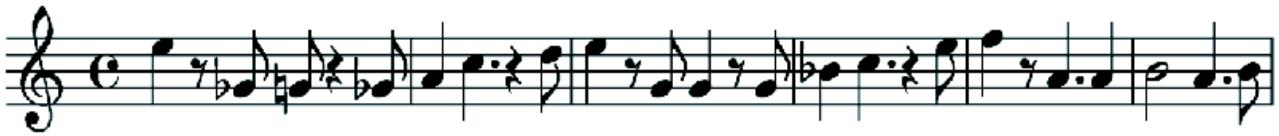
Pitch and tempo invariance note encoding.



Melody Matching

Coding

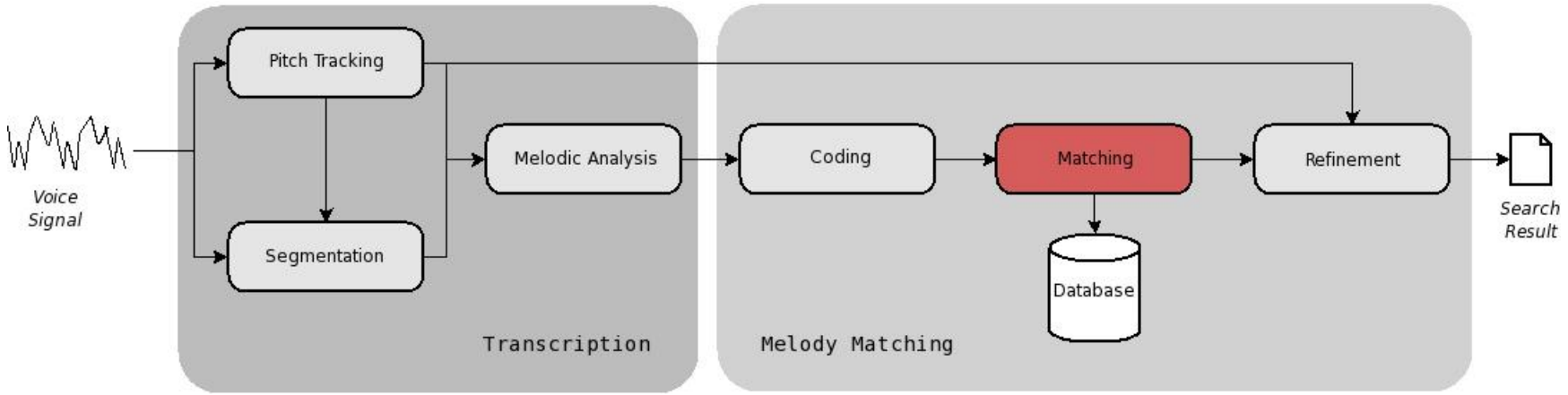
- Pitch Intervals
  $A = (a_1, a_2, \ldots, a_n) \rightarrow \overline{A} = (a_2 - a_1, a_3 - a_2, \ldots, a_n - a_{n-1})$
- Relative Durations
  $D = (d_1, d_2, \ldots, d_n) \rightarrow \overline{D} = (\frac{d_2}{d_1}, \frac{d_3}{d_2}, \ldots, \frac{d_n}{d_{n-1}})$



MIDI Note	53	43	44	43	46	49	51	53	44	44	44	47	49	53	54	46	46	48	46	48
Pitch Interval	*	-10	1	-1	3	3	2	2	-9	0	0	3	2	4	1	-8	0	2	-2	2
Duration (♪)	3	1	3	1	2	5	1	3	1	3	1	2	5	1	3	3	2	4	3	1
Relative Duration	*	$\frac{1}{3}$	3	$\frac{1}{3}$	2	$\frac{5}{2}$	$\frac{1}{5}$	3	$\frac{1}{3}$	3	$\frac{1}{3}$	2	$\frac{5}{2}$	$\frac{1}{5}$	3	1	$\frac{2}{3}$	2	$\frac{3}{4}$	$\frac{1}{3}$

Matching

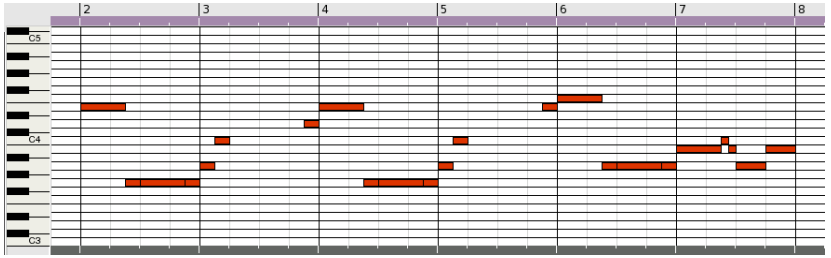
Find good occurences of the codified query in the database.

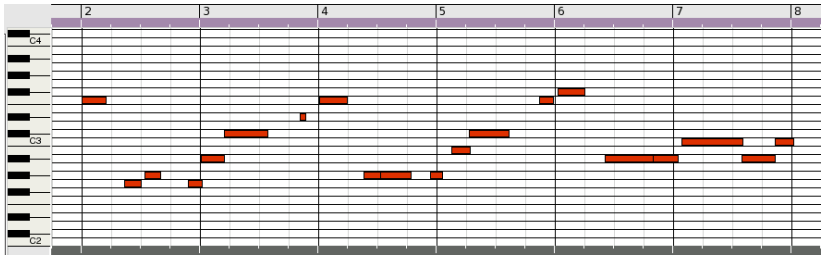


Note by Note Matching

String Matching

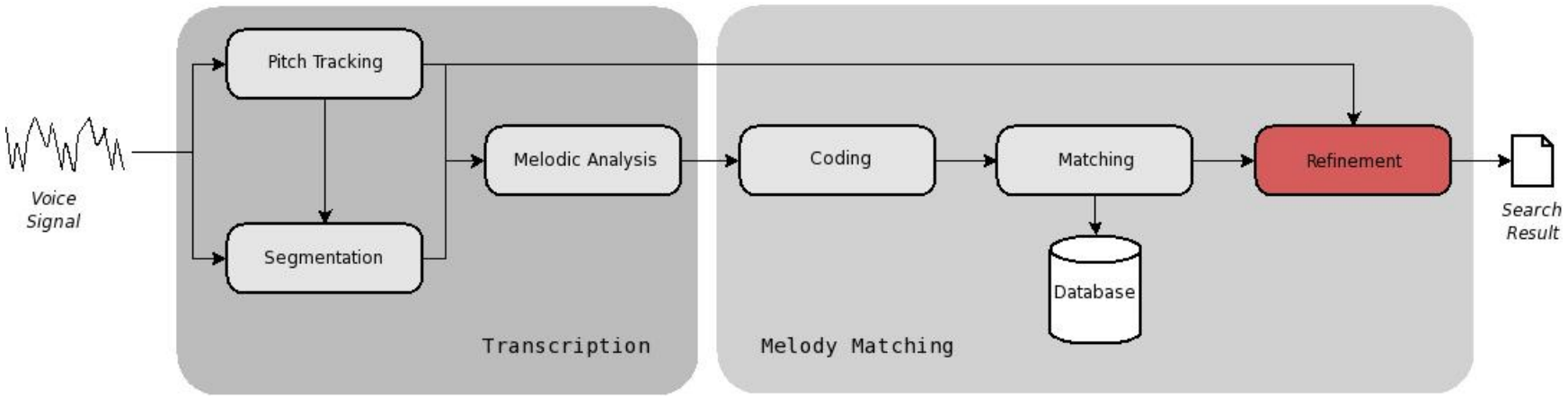
- Edit Distance combining pitch and duration
- Calculated using Dynamic Programming





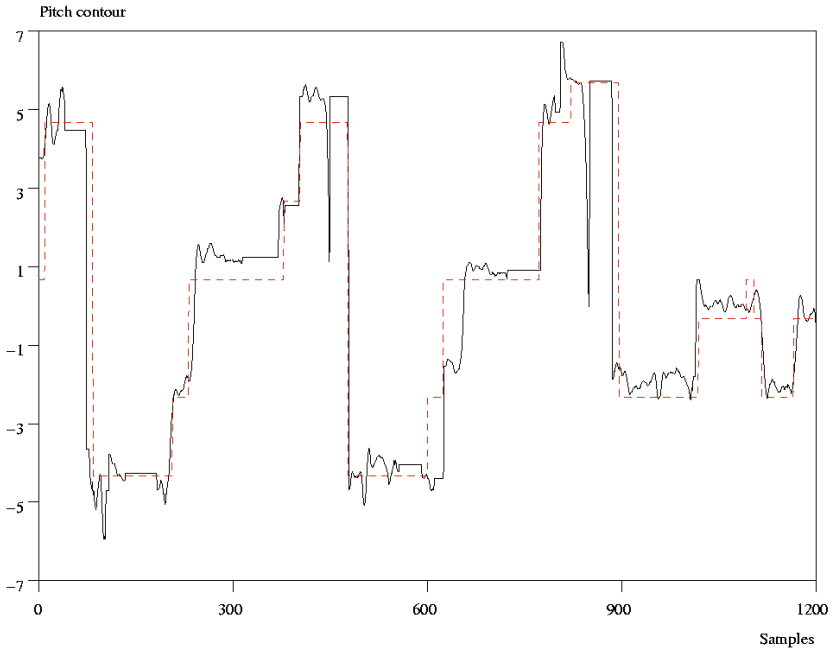
Pitch Time Series Matching

Refine candidates selection by pitch time series comparison.



Pitch Time Series Matching

Local Dynamic Time Warping



Note by note matching enables:

- Reduced candidates group
- Similar melody fragments identification

LDTW restrictions are avoided.

Outline

Evaluation

- MIDI Database with The Beatles collection (208 songs)
- More than 30 untrained subjects took part
- Top 1 overall performance: 75%

	Singing		**Humming**	
	Short	Long	Short	Long
Number of files	68	242	85	32
Note average	11.15	25.81	12.03	22.40
Top 1 (%)	73.53	76.86	75.00	83.52
Top 10 (%)	80.88	83.06	90.62	89.41

Outline

Conclusions

Development Requirements

A QBH system must:

- be tolerant to query errors
- be unrestictive in the way of singing
- retrieve only the desired piece
- allow easy database extension

Contributions

- Sufficiently robust transcription system
- Novel way of combining matching techniques

de Cheveigné, A. y Kawahara, H.
Yin, a fundamental frequency estimator for speech and music.
JASA, 111:1917–1930, 2002.

Klapuri, A.P.
Sound onset detection by applying psichoacoustic knowledge.
ICASSP, 1999.

Pollastri, E. and Haus G.
An audio front end for query-by-humming systems.
Proc. of ISMIR, 2001.

Tararira is Free Software and can be downloaded from:
`http://iie.fing.edu.uy/investigacion/grupos/gmm/proyectos/tararira/`